干旱区地理
ARID LAND GEOGRAPHY

基于多光谱遥感的典型绿洲棉田春季土壤盐分反演及验证

刘旭辉[1,2]， 白云岗[2]， 柴仲平[1]， 张江辉[2]， 丁邦新[2,3]， 江 柱[2]

（1. 新疆农业大学资源与环境学院，新疆 乌鲁木齐 830052；2. 新疆水利水电科学研究院，新疆 乌鲁木齐 830049；3. 西北农林科技大学水利与建筑工程学院，陕西 杨凌 712100）

摘 要：为探索快速提取典型绿洲棉田土壤盐分的有效方法，获取区域尺度的土壤盐渍化特征及空间分布，进而为土壤盐渍化防治提供参考。以新疆兵团农二师31团为研究区域，2019、2021年春季Landsat 8 OLI多光谱影像和野外实测土壤含盐量为数据源，将波段组、光谱指数组和全变量组作为模型输入变量组，采用多元逐步回归（Multiple stepwise regression, MSR）、偏最小二乘回归（Partial least squares regression, PLSR）、极限学习机（Extreme learning machine, ELM）、支持向量机（Support vector machine, SVM）和BP神经网络（Back propagation neural network, BPNN）构建基于3个输入变量组的土壤盐分遥感反演模型，探究输入变量和建模方法对模型精度的影响效果，通过对比确定春季土壤盐分最优反演模型，定量反演地表土壤含盐量。结果表明：（1）研究区主要为非盐化土和轻度盐化土，总样本变异系数为0.67，呈中等变异性；光谱反射率与土壤盐渍化程度的关系表现为土壤盐渍化越重，光谱反射率越高。（2）海岸波段（b1）、蓝波段（b2）、绿波段（b3）、红波段（b4）和盐分指数（SI1、SI2、SI3、SI4、S3、S4、S5）均通过显著性检验$P<0.01$，相关系数均达到0.4以上。（3）所有模型中，基于全变量组建立的BPNN反演模型精度最高，建模集R^2为0.705；验证集R^2为0.556。（4）由反演结果可知，2019、2021年春季耕作区土壤主要为非盐化土，分别占耕作区总面积的55.55%和64.62%，其次为轻度盐化土，分别占44.31%和35.17%；2021年土壤盐渍化程度较2019年有所减轻。

关 键 词：多光谱遥感反演；土壤盐分；光谱反射率；变量组；机器学习

文章编号：

目前，土壤盐渍化已经成为全球性生态环境问题，严重制约着干旱区灌溉农业的可持续发展[1-2]。新疆作为我国重要的农垦区，盐渍土面积约占耕地总面积的1/3，盐渍化问题尤为严重[3]。因此，亟需快速准确获取土壤含盐量变化的监测方法，以支持灌区灌溉制度制定、土壤盐渍化防治、盐碱地综合利用、土地管理等工作[4]。传统的实地采样调查方法不仅耗费大量的人力、物力，而且研究尺度较小，获取的多为点状信息[5-6]。遥感技术能够大面积快速获取同一地区不同时间点的地物信息，有助于开展长时间序列的土壤盐渍化研究工作，加之宏观、综合、及时、动态、低成本等特点，被广泛应用于土壤盐渍化的实时监测和土壤盐分的定量估算等[7]。

随着遥感技术的愈加成熟，学者们开始深入探索提高建模精度的各类方法，其中，光谱参量的构建和建模方法的选取有助于提高遥感反演精度[8]。许多学者采用敏感波段组合、特征光谱指数构建和光谱数学变换等方法创建模型输入变量[9-10]。周晓红等[10]以艾比湖湿地自然保护区为试验区，基于Landsat 8 OLI多光谱遥感影像，构建了多种植被指数和盐分指数，在众多光谱指数中，增强型植被指数与土壤盐分的相关性最高，传统型植被指数次之，盐分指数最低；张晓光等[11]对光谱数据分别进行了对数、倒数、一阶微分、二阶微分等数学变换，以增强光谱与盐分之间的相关性，进而达到了提升

收稿日期：2021-10-15； **修订日期**：2022-01-01

作者简介：刘旭辉（1996-），女，硕士研究生，主要从事干旱区资源与环境遥感研究. E-mail: 454949752@qq.com

通讯作者：白云岗（1974-），男，博士，教授级高级工程师，主要从事农业水土工程研究. E-mail: xjbaiyg@sina.com

模型反演精度的效果;Zhang等[12]基于MODIS卫星影像构建了光谱指数,反演了黄河三角洲地区的土壤含盐量;Alexakis等[13]针对希腊克里特岛土壤盐渍化问题,基于WV2和Landsat 8 OLI计算盐分指数,结果表明,光谱指数S5对土壤盐分最为敏感,可为该地区地表土壤含盐量的监测提供新方案。以上研究表明,筛选敏感光谱变量可以有效提升模型反演效果[8]。除了自变量的优化,建模方法也经历了一段发展历程,从传统的线性回归发展到具有非线性拟合能力的机器学习算法。研究发现,当模型输入变量个数较多、且与土壤盐分的相关性较低时,机器学习的优势得到了更加充分的体现[14]。通过对比不同模型的建模和验证效果,确定了最适合本研究区土壤盐分的建模方法和反演模型,进而提升了土壤盐分遥感反演精度。冯雪力等[15]对比了传统回归方法和机器学习算法的建模精度,发现BP神经网络(Back propagation neural network, BPNN)模型预测效果最优,其稳定性和预测精度均优于经验回归模型;刘恩等[16]基于Landsat 8 OLI多光谱影像,分别采用多元线性回归和BP神经网络算法构建了土壤盐分反演模型,发现影像光谱反射率和土壤含盐量之间不是单纯的线性关系,因而多元回归模型的预测效果具有很大的波动性,而BPNN具有极强的非线性拟合能力,因而机器学习模型反演精度更高;张智韬等[17]以内蒙古河套灌区为研究区域,采用多种方法进行建模,发现支持向量机(Support vector machine, SVM)模型精度最高,分位数回归次之,偏最小二乘回归(Partial least squares regression, PLSR)最低。以上研究都说明了机器学习算法在土壤盐分遥感反演过程中的优越性。

土壤盐渍化相关研究虽然取得了一系列较为理想的成果,但现有的遥感反演研究多是基于某一特定时期的遥感影像展开,仅能反映当地该时期的土壤盐渍化状况,模型的推广性不强。而且,盐渍化程度受时空变化影响,不同位置的土壤含盐量存在差异性。因此,探讨某一地区不同年份的土壤盐渍化状况极为重要[18-19]。鉴于此,本文以新疆兵团农二师31团为研究区域,基于2019、2021年的Landsat 8 OLI卫星遥感影像和相应时期地表土壤含盐量,利用多元逐步回归(Multiple stepwise regression, MSR)、PLSR、极限学习机(Extreme learning machine, ELM)、SVM和BPNN构建土壤盐分遥感反演模型,实现棉田土壤盐分的定量反演,以期为当地土壤表层含盐量的实时监测提供参考。

1 材料与方法

1.1 研究区概况

新疆生产建设兵团农二师31团位于天山南麓塔里木盆地东北边缘,塔里木河与孔雀河两河下游的冲积平原上,地处巴音郭楞蒙古自治州尉犁县境内,西南临塔克拉玛干沙漠,东北与库姆塔格沙漠接壤,地理坐标为85°24′~88°30′E,39°30′~42°20′N。灌区土壤类型主要为盐土、碱土和风沙土。团场深处欧亚内陆腹地,属北温带大陆性荒漠干旱气候,降水量年际变化大,多年平均降水量53.3~62.7 mm,多年平均蒸发量2273~2788 mm。现有灌溉面积7.5×10^3 hm^2,灌溉方式以膜下滴灌为主,灌溉用水来自塔里木河与恰拉水库。生育期灌溉量2700~3450 m^3·hm^{-2},自每年5月开始;冬灌量约为3600 m^3·hm^{-2},自每年11月开始,使用大水漫灌。地下水位保持在1~3 m之间,地下水矿化度2~7 g·L^{-1};塔里木河与水库矿化度0.47~1.53 g·L^{-1};灌溉排水矿化度10.58~18.36 g·L^{-1}。

1.2 野外数据采集与处理

野外采样选择在棉花播种前(2019年3月20—28日、2021年4月7—12日)进行,该时段为裸土期,受植被影响较小。实地勘察后,按照从水库引水处沿干渠走向布设采样点,同时考虑交通条件、分布位置、土壤类型等因素。2019年和2021年最终采集的样点个数分别为53个和33个(图1)。

1.2.1 *表层土壤含盐量测定* 考虑到土壤水盐的变异性较强,从预设点向外辐射约10 m的范围内再选取2个点,利用3点取样法采集地表0~20 cm深度的土壤,使用四分法将3处土壤混合,作为该编号位置上的一个样品,随即将样品装入已编号的自封袋中,手持GPS仪测定每个样点的经纬度,拍照记录各样点处的盐渍化程度、作物类型、植被覆盖度等信息。将土样带回实验室后,挑出杂物,风干,研磨,过2 mm孔径筛,称取18 g土样,在锥形瓶中按土水质量比为1:5配置土壤溶液,用玻璃棒搅拌均匀后固定在振荡器上震荡10 min,使盐分充分溶解,静置后使用电导率仪测定土壤悬浊液的电导率值(EC,μS·cm^{-1}),根据经验公式:$y=0.0051x-0.5241$(决

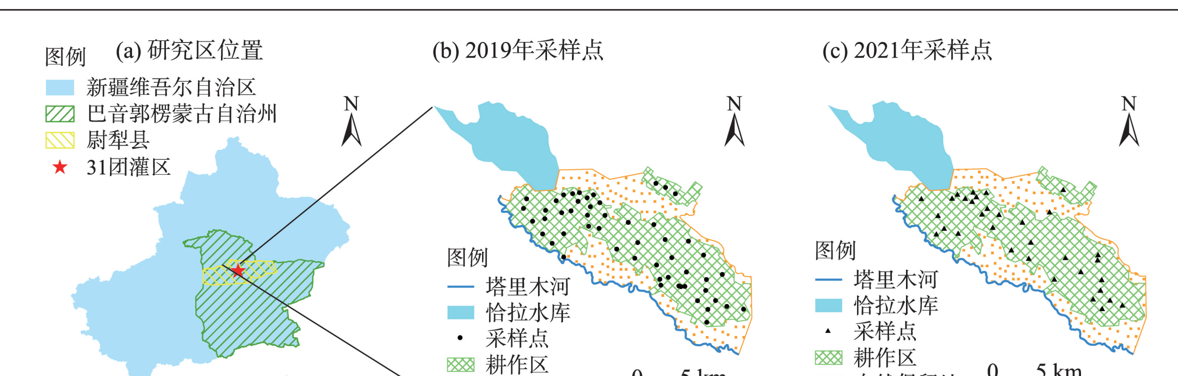


图1　研究区位置及土壤样点分布示意图

Fig. 1　Sketch maps of the study area location and soil sample distribution

定系数 R^2=0.9534)计算土壤总盐($g\cdot kg^{-1}$)[20]。

1.2.2　*表层土壤光谱曲线采集与处理*　选择晴朗、无云、无风的日子,在12:00—16:00之间进行光谱测定。使用美国PP Systems公司生产的UniSpec-SC光谱分析仪(波段范围310~1130 nm)测定土壤光谱反射率。样点观测前需要进行标准白板扫描和暗扫描操作,光谱采集时,探头设置在垂直距离地面50 cm处,从而缩小仪器视场角,以减小地面背景及周围其他地物对土壤光谱的影响。对不同盐渍化等级的土壤测定多条光谱曲线,每个等级测定10次,取平均值后作为每个盐渍化等级土壤的实测高光谱数据。土壤光谱采集完毕后,使用Multispec 5.1.5软件对数据进行处理,得到310~1130 nm波段范围的土壤光谱反射率,在Origin 2018软件中绘制土壤实测高光谱反射率曲线。

1.3　遥感影像数据的获取与处理

多光谱数据采用Landsat 8 OLI卫星遥感影像,数据来源于美国地质调查局(United States Geological Survey, USGS)官网(https://earthexplorer.usgs.gov/),行列号143/031,空间分辨率30 m,重访周期16 d。根据与采样时间相近且云量较小的原则,本文选取2019年3月22日和2021年3月20日的影像作为本次试验的多光谱数据源。使用ENVI 5.3软件对影像数据进行辐射定标、大气校正和裁剪等预处理操作,实现遥感数据由数字量化值到辐射亮度值再到地表反射率的逐步转换,再通过ArcGIS 10.2提取各采样点对应的影像多光谱反射率,在Excel 2016中计算各类型光谱指数。

1.4　光谱参量构建

本文选取了7个波段和12个盐分指数作为土壤盐分反演模型的备选输入变量,光谱参量及计算公式如表1所示,包括海岸波段(b1)、蓝波段(b2)、绿波段(b3)、红波段(b4)、近红外波段(b5)、短波红外1波段(b6)、短波红外2波段(b7)、盐分指数[归一化盐分指数(NDSI)、SI-T、SI1、SI2、SI3、SI4、S1、S2、S3、S4、S5、S6]。

表1　光谱参量及相关计算公式

Tab. 1　Spectral parameters and related formulas

光谱参量		计算公式	参考文献
波段	b1	CA	[16]
	b2	B	[16]
	b3	G	[16]
	b4	R	[16]
	b5	NIR	[16]
	b6	SWIR1	[16]
	b7	SWIR2	[16]
盐分指数	NDSI	$\frac{R-\mathrm{NIR}}{R+\mathrm{NIR}}$	[14]
	SI-T	$100\times(R-\mathrm{NIR})$	[14]
	SI1	$\sqrt{G\times R}$	[2]
	SI2	$\sqrt{G^2+R^2+\mathrm{NIR}^2}$	[2]
	SI3	$\sqrt{G^2+R^2}$	[2]
	SI4	$\sqrt{(B-G)^2+(G-R)^2}$	[10]
	S1	B/R	[2]
	S2	$(B-R)/(B+R)$	[2]
	S3	$(G\times R)/B$	[2]
	S4	$\sqrt{B\times R}$	[2]
	S5	$(B\times R)/G$	[2]
	S6	$(R\times \mathrm{NIR})/G$	[2]

注:CA、B、G、R、NIR、SWIR1、SWIR2分别为海岸波段(b1)、蓝波段(b2)、绿波段(b3)、红波段(b4)、近红外波段(b5)、短波红外1波段(b6)、短波红外2波段(b7)的光谱反射率(%);NDSI为归一化盐分指数;SI-T、SI1、SI2、SI3、SI4、S1、S2、S3、S4、S5、S6均为盐分指数。下同。

1.5 遥感反演模型构建

考虑到土壤盐分受水分的影响，不同时期的灌溉、气温和降水等因素会影响土壤水分和盐分含量，2期土壤含盐量分布的变异性存在差异。剔除异常值后，将2019年和2021年春季样本数据混合，按照土壤含盐量由小到大的顺序排列分组，根据建模集与验证集2:1的比例进行等间隔取样[14]，最终确定了建模集样本51个，验证集样本26个。对建模集影像的波段、盐分指数与土壤含盐量进行皮尔逊相关性分析(Pearson correlation coefficient, PCC)，通过分析进行特征光谱参量筛选，将筛出的特征光谱参量分为波段组、光谱指数组和全变量组，并作为输入变量组参与建模，建模方法选用MSR、PLSR、ELM、SVM和BPNN，共计15个土壤盐分反演模型。建模方法如下：

(1) 多元逐步回归(MSR)

MSR在多元线性回归的基础上进行改进，它能进一步优化公式，确定各个自变量对于因变量预测值的比重，剔除不重要的变量[15,21]，是土壤盐分遥感反演最常用的回归方法之一[22]。本研究的MSR建模与分析基于IBM SPSS Statistics 25完成。

(2) 偏最小二乘回归(PLSR)

PLSR是一种多因变量对多自变量的回归建模方法，能够在一定程度上消除参量之间的多重共线性，通过不断地迭代，最终模型将包含原有的所有自变量[15,23]。本文采用The Unscrambler X 10.4软件进行建模与分析。

(3) 极限学习机(ELM)

ELM是一种单隐层前馈型神经网络，具有泛化能力强、人工干预小、学习速度快等优势[24]，其特点是输入层与隐含层之间的连接权值和隐含层神经元的阈值随机生成[25]。本研究基于MATLAB 2018平台进行ELM模型构建，通过反复试验，最终确定了波段组的神经网络模型结构(输入层神经元个数-隐含层神经元个数-输出层神经元个数)为4-11-1，光谱指数组的结构为7-11-1，全变量组为11-7-1。

(4) 支持向量机(SVM)

SVM的原理是建立一个最优分类超平面，使其分类间隔最大化[26]。在进行非线性问题的求解时具有泛化能力强、全局最优、结构化风险最小等优势[27]。本文采用LIBSVM工具箱在MATLAB 2018软件中建立SVM模型，通过测试确定了最适于本实验数据的svmtrain函数参数，SVM类型设置为e-SVR回归，核函数选用高斯径向基核函数(RBF)。惩罚参数(c)和核函数参数(g)根据样本数据进行测试，最终确定了波段组c为900，g为0.25；光谱指数组c为367，g为0.14；全变量组c为4，g为0.09。

(5) BP神经网络(BPNN)

BP神经网络(多层前馈式误差反向传播神经网络)是一种监督学习模型，它通过对代表性样本进行学习训练，从而掌握研究对象的本质特性，具有很强的自组织和自适应能力，能够模拟任意的非线性输入输出关系[28]。本研究基于MATLAB 2018平台，进行BPNN的结构设计。将建模集的波段组、光谱指数组和全变量组分别作为模型输入因子，土壤盐分作为模型输出因子，建立土壤盐分BPNN反演模型。经过反复训练，最终确定了最适于本实验数据的BPNN模型参数，训练函数选用poslin，隐含层传递函数选用purelin，输出层函数选用trainlm，模型最大迭代次数1000次，训练目标最小误差0.00001，波段组的神经网络结构为4-9-1，光谱指数组的结构为7-9-1，全变量组为11-9-1。

1.6 模型评价指标

为量化模型的拟合能力，评价指标采用决定系数(Coefficient of determination, R^2)和均方根误差(Root mean square error, RMSE)[29]对模型的建模和验证效果进行综合评价。R^2越接近1，表示模型拟合程度越好；RMSE越小，表示实测值和预测值误差越小。因此，R^2越大，RMSE越小，说明模型精度越高，反演效果越好。

2 结果与分析

2.1 土壤盐分特征

根据新疆土壤盐碱化程度等级分类标准[30]，将土壤分为非盐化土(0~3 $g\cdot kg^{-1}$)、轻度盐化土(3~6 $g\cdot kg^{-1}$)和中度盐化土(6~10 $g\cdot kg^{-1}$)3个等级。土壤含盐量的分析结果如表2所示，总样本统计结果显示，非盐化土、轻度盐化土、中度盐化土所占比例分别为64.94%、27.27%、7.79%，变异系数为0.67，表明棉田土壤含盐量呈中等变异性。

2.2 实测高光谱特征

如图2所示，土壤高光谱波长范围为310~1130

表2 土壤盐分的描述性统计分析

Tab. 2 Descriptive statistical analysis on soil salinity

数据集	样本数/个				土壤含盐量/g·kg^{-1}			标准差	变异系数
	总样本	非盐化土	轻度盐化土	中度盐化土	最小值	最大值	平均值		
总样本	77	50	21	6	0.37	11.82	3.02	2.03	0.67
建模集	51	33	14	4	0.42	11.82	3.05	2.07	0.68
验证集	26	17	7	2	0.37	9.64	2.97	1.95	0.66

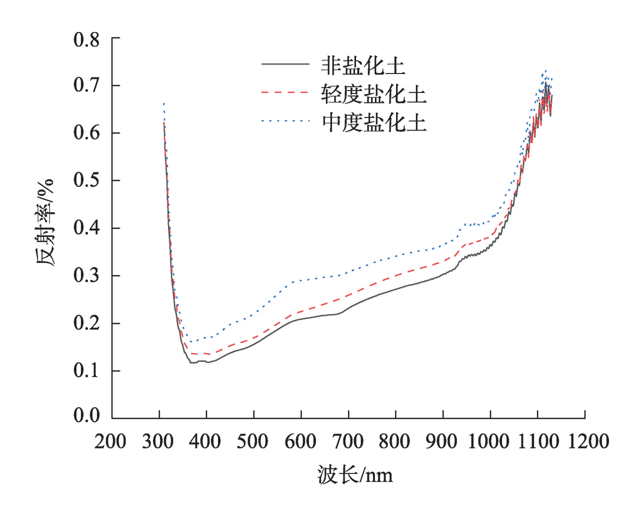


图2 不同盐渍化等级土壤光谱反射率曲线

Fig. 2 Spectral reflectance curves of soil with different salinization grades

nm。地表土壤含盐量不同导致土壤反射率存在差异，但不同盐渍化等级土壤的光谱曲线相对平行，形态相似，变化趋势基本一致，随着波长的增加，土壤高光谱反射率先减小后增大。波长在310~373 nm之间，光谱反射率随波长的增加而迅速减小；波长在373~1130 nm之间，反射率随波长的增加而增大，其中在373~1050 nm之间，曲线增长较缓，当大于1050 nm时，反射率迅速增大。整体来看，土壤盐渍化越重，光谱反射率越高。

2.3 特征光谱参量优选

根据皮尔逊相关系数界值表[14]，样本数(n)等于50时，当相关系数(r)的绝对值大于0.354时，表示在0.01水平上显著，光谱参量与土壤含盐量的Pearson相关系数如表3所示，筛选出敏感波段(P<0.01)和敏感光谱指数(P<0.01)。b1、b2、b3、b4波段显著性检验P<0.01，且r均达到0.5以上，与土壤含盐量的相关性大小依次为b1>b2>b3>b4。SI1、SI2、SI3、SI4、S3、S4、S5盐分指数与土壤含盐量均呈显著相关(P<0.01)，相关性大小依次为S4>S5>SI1>SI3>S3>SI2>SI4，且r均达到0.4以上，说明SI1、SI2、SI3、SI4、S3、S4、S5在一定程度上可以表征土壤含盐量。

表3 光谱参量与土壤盐分的相关性

Tab. 3 Correlation between spectral parameters and soil salinity

光谱参量		相关系数(r)
波段	b1	0.591**
	b2	0.550**
	b3	0.539**
	b4	0.525**
	b5	0.297
	b6	0.196
	b7	0.329
盐分指数	NDSI	0.304
	SI-T	0.197
	SI1	0.535**
	SI2	0.452**
	SI3	0.533**
	SI4	0.416**
	S1	−0.162
	S2	−0.163
	S3	0.497**
	S4	0.545**
	S5	0.540**
	S6	0.269

注：**表示显著性检验P<0.01。下同。

2.4 土壤盐分反演模型构建

选择与土壤盐分显著相关(P<0.01)的b1、b2、b3、b4波段作为波段组，选择与土壤盐分显著相关(P<0.01)的SI1、SI2、SI3、SI4、S3、S4、S5盐分指数作为光谱指数组。为了避免模型中遗漏重要的光谱参量，且考虑不同类别光谱参量对模型的影响，将上述4个敏感波段和7个敏感光谱指数共同作为全变量组参与建模。

2.4.1 土壤盐分回归模型 由表4可知，同一变量组作为模型输入变量下，MSR模型建模集R^2均高于PLSR，且数值均大于0.340，RMSE均低于PLSR，且数值均小于1.680，说明MSR建模效果较好。对模

表4 基于不同变量组的回归模型

Tab. 4 Regression models based on different groups of variables

变量组	建模方法	表达式	建模集(n=51)		验证集(n=26)	
			R^2	RMSE	R^2	RMSE
波段组	MSR	Y=624.08b1−440.71b2−29.42	0.411	1.588	0.473	1.414
	PLSR	Y=40.75b1+35.51b2+25.32b3+19.94b4−21.19	0.314	1.713	0.496	1.382
光谱指数组	MSR	Y=−94.05SI2+170.77SI4+246.21S5−14.30	0.341	1.679	0.377	1.544
	PLSR	Y=13.32SI1+6.02SI2+9.3SI3+26.55SI4+7.86S3+16S4+16.58S5−16.74	0.269	1.768	0.449	1.447
全变量组	MSR	Y=592.81b1−318.31b3−75.19SI2+462.30SI4−23.33	0.439	1.549	0.359	1.562
	PLSR	Y=15.29b1+13.33b2+9.5b3+7.48b4+8.53SI1+3.86SI2+5.96SI3+17.01SI4+5.03S3+10.25S4+10.62S5−18.73	0.290	1.742	0.470	1.418

注:MSR为多元逐步回归;PLSR为偏最小二乘回归;R^2为决定系数;RMSE为均方根误差。下同。

型进行精度验证时,发现波段组的MSR模型和PLSR模型R^2相近,均大于0.472,RMSE均小于1.415,验证效果较好。在光谱指数组和全变量组中,PLSR模型验证集R^2均高于MSR,其数值均大于0.448,RMSE均低于MSR,其数值均小于1.448,说明PLSR验证效果较好。综合对比两种回归方法的不同变量组建模和验证效果,发现基于波段组建立的MSR模型是表4所有模型中最优的土壤盐分反演模型。

2.4.2 土壤盐分机器学习算法反演模型 如表5所示,波段组和光谱指数组作为模型输入变量下,建模集的各类机器学习模型R^2均大于0.616,其中SVM模型的建模集R^2均达到0.768以上,RMSE均小于0.998,建模精度最高,BPNN模型次之,ELM模型精度最低。对模型进行精度验证时,发现波段组和光谱指数组的ELM、SVM模型RMSE较大。在全变量组中,BPNN模型的建模集和验证集R^2均为3个模型中的最高值,分别达到了0.705和0.556,RMSE分别为1.133和1.409,模型精度最高,ELM模型次之,SVM模型精度最低。因此,在3种机器学习算法中,BPNN模型最适于本实验数据。综合对比3种算法的不同变量组建模和验证效果,发现基于全变量组建立的BPNN模型是表5所有模型中最优的土壤盐分反演模型。对比表4和表5可知,机器学习模型精度均高于线性回归模型,因此,本文最终使用基于全变量组建立的BPNN模型来反演31团春季土壤表层含盐量。

2.5 土壤盐分空间分布反演

对比图3和图4可知,土壤含盐量为6 g·kg^{-1}以上的土壤均分布于灌区中部的盐碱地带上,含盐量为0~3 g·kg^{-1}的土壤均分布于灌区内部远离荒地的区域,越靠近荒地,土壤含盐量越高。由此可知,土壤盐分反演结果与Kriging插值结果基本一致,与调查情况相符,表明耕地盐分反演结果的准确率较高。

如表6所示,2019、2021年春季耕作区土壤主要为非盐化土,分别占耕作区总面积的55.55%和

表5 基于不同变量组的机器学习模型

Tab. 5 Machine learning models based on different variable groups

变量组	建模方法	建模集(n=51)		验证集(n=26)	
		R^2	RMSE	R^2	RMSE
波段组	ELM	0.617	1.279	0.533	1.998
	SVM	0.782	0.985	0.476	1.996
	BPNN	0.736	1.067	0.538	1.460
光谱指数组	ELM	0.621	1.274	0.508	1.986
	SVM	0.769	0.997	0.409	2.001
	BPNN	0.730	1.080	0.459	1.567
全变量组	ELM	0.540	1.403	0.490	1.565
	SVM	0.533	1.515	0.488	1.658
	BPNN	0.705	1.133	0.556	1.409

注:ELM、SVM和BPNN分别为极限学习机、支持向量机和BP神经网络。

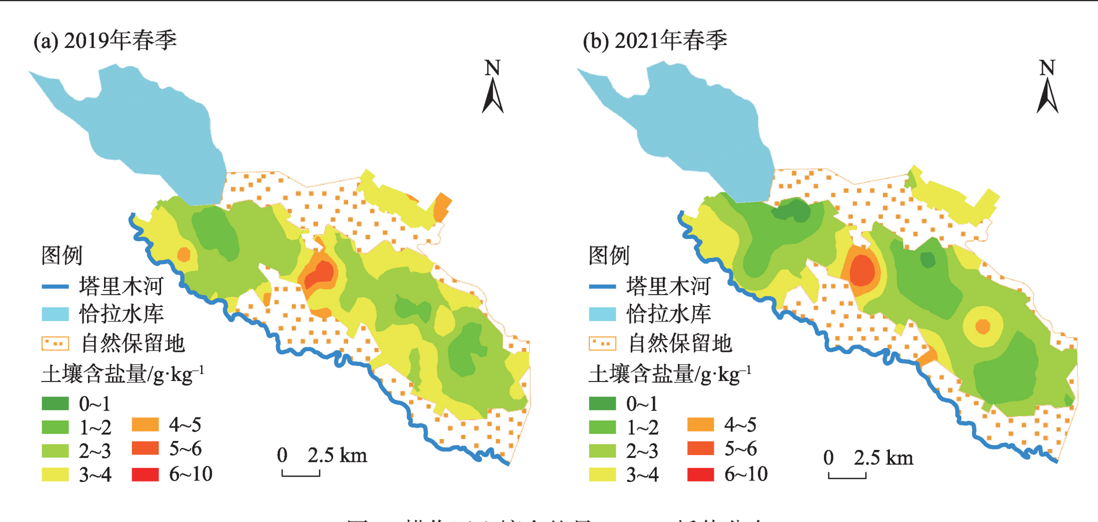


图3 耕作区土壤含盐量Kriging插值分布

Fig. 3 Kriging interpolation distributions of soil salinity in tillage area

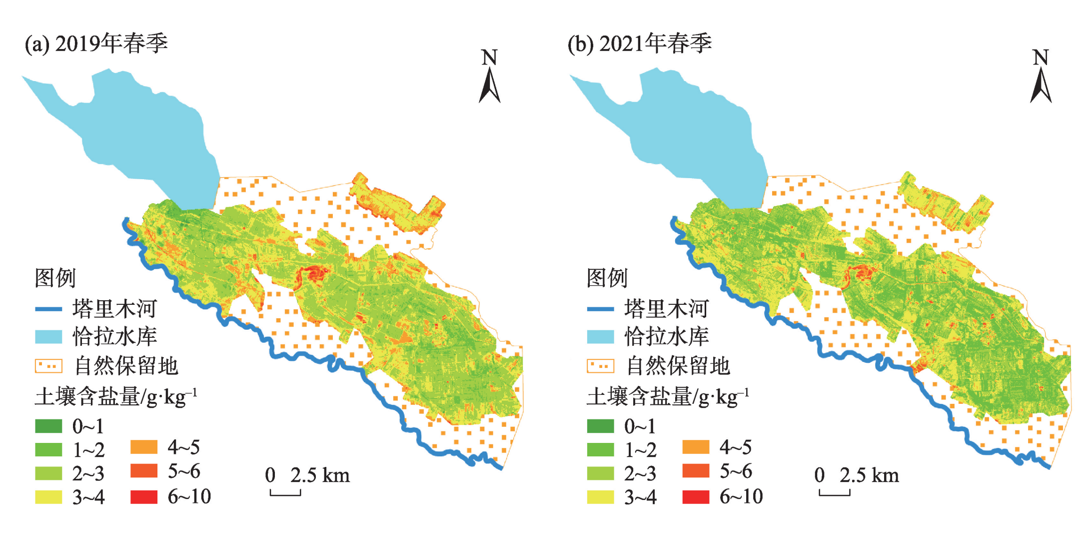


图4 耕作区土壤含盐量反演等级分布

Fig. 4 Inversion grade distributions of soil salinity in tillage area

表6 不同盐分层次的像元个数及比例

Tab. 6 Number and proportion of pixels at different salinity levels

土壤类型	土壤含盐量/g·kg^{-1}	2019年		2021年	
		像元个数	面积占比/%	像元个数	面积占比/%
非盐化土(0~3 g·kg^{-1})	0~1	4508	0.35	14	0.00
	1~2	92995	7.17	366240	28.25
	2~3	622671	48.03	471548	36.37
轻度盐化土(3~6 g·kg^{-1})	3~4	428918	33.08	398125	30.71
	4~5	128548	9.92	47257	3.65
	5~6	16989	1.31	10556	0.81
中度盐化土(6~10 g·kg^{-1})	6~10	1848	0.14	2737	0.21

64.62%，其中含盐量为2~3 g·kg^{-1}的土壤分布面积最大，1~2 g·kg^{-1}的土壤分布面积次之，0~1 g·kg^{-1}的土壤分布面积最小。轻度盐化土面积分别占44.31%和35.17%，其中含盐量为3~4 g·kg^{-1}的土壤分布面积最大，4~5 g·kg^{-1}的土壤分布面积次之，5~6 g·kg^{-1}的土壤分布面积最小。对比2 a土壤盐渍化

程度,发现2021年春季非盐化土面积相比2019年增加了9.07%,轻度盐化土面积缩小了9.14%,中度盐化土面积变化较小。

由图4可知,含盐量小于2 g·kg⁻¹的土壤主要分布于恰拉水库附近、耕作区中部和东南部。越接近南北两侧荒地的耕地土壤含盐量越高,东北部的小块耕地盐分偏高,土壤盐渍化程度主要为轻度盐化土。中度盐化土主要分布于灌区中部的盐渍化地带上,与实际影像位置相重合,取得了较为准确的反演效果。此外,东北部的小块耕地部分区域和灌区内的零散地块上也有少量分布。2021年反演图中绿色更深,即含盐量小于2 g·kg⁻¹的土壤分布面积更大,在整个灌区都有体现。2021年含盐量大于4 g·kg⁻¹的土壤分布面积相比于2019年有所减小,图上表现为橙红色区域缩小。综上可知,2021年土壤盐渍化程度较2019年有所减轻。

3 讨 论

3.1 盐渍土特征分析及光谱参量优选

棉田土壤含盐量呈中等变异性,说明其易受气候、地形、水文、地质等自然因素和灌溉、翻耕、施肥等人为因素的影响[31-33]。采样时间为2019年和2021年春季,2 a自然环境差异较大,不可控因素较多,因而本文将2 a春季数据混合,以中和不同年份的样本数据变异性,再进行建模集和验证集划分,更具有合理性。本文对建模集OLI多光谱波段、光谱指数与实测土壤盐分进行相关性分析,发现多光谱敏感波段为蓝、绿、红波段,这与孙亚楠等[2]研究结果一致,说明波段为430~760 nm的土壤盐分光谱信息量较多,而通过波段组进行特征光谱指数的构建可以挖掘光谱与盐分之间的隐含信息,推动土壤盐分遥感定量反演研究的发展。王雪梅等[34]、边玲玲等[35]发现光谱指数与土壤盐分之间存在一定的相关性,在估算应用方面具有很大贡献。本研究也显示,当敏感波段与敏感光谱指数共同作为模型输入变量时,土壤盐分反演模型的建模和验证效果最优。

3.2 土壤盐分反演模型建立

本文基于3个变量组,使用5种方法建立了15个土壤盐分反演模型,发现选择不同的输入变量和建模方法,模型效果存在差异。研究发现,基于机器学习模型的精度和稳定性均优于传统线性回归,这是由于土壤盐渍化发生机理的复杂性,土壤光谱特征与表层土壤含盐量之间存在复杂的非线性关系,而机器学习算法具有极强的非线性拟合能力和推广能力,适合模拟变量内部机制的复杂关系。许多学者研究发现机器学习算法的反演效果优于传统回归模型[14,23],与本文的研究结果一致。冯雪力等[15]在进行PCR、MSR和PLSR建模时,发现MSR模型模拟全盐量效果最优,本研究也发现基于波段组建立的MSR模型是6个线性回归模型中最优的土壤盐分反演模型。传统线性回归建模效果虽次于机器学习算法,但其简便易懂,回归方程清楚可见,具有一定的运用价值;机器学习模型可以极大的提升模型反演精度,也具有一定的研究价值。

3.3 土壤盐分时空变化特征

土壤盐分Kriging插值图基于各采样点处土壤盐分的实测值生成,数据源为点状数据,数据量较小,成果图较为粗略,仅能大致反应土壤盐分的分布状况。而遥感技术能根据土壤盐分反演模型将影像上所有栅格点处的光谱反射率转化为土壤含盐量,由此达到量化效果,再对栅格点数值进行范围划分,可以得出不同盐分层次的栅格点个数及所占比例。对比2 a土壤盐渍化程度,发现2021年土壤盐渍化程度较2019年有所减弱,这可能与气温、蒸发量、灌水量有关。气温越高,蒸发作用越强,土壤中的盐分向表层运移;灌水量越大,对土体淋洗作用越强,土壤表层含盐量越低。而且当地非常重视盐渍土的治理,制定了合理的农业灌溉排水制度和措施,随着时间的推移,取得了一定成效。

已有研究多是针对某一年份的盐渍土进行反演[36-37],但由于每年环境气候多变,且栅格点内地物类型复杂,使得不同年份的遥感影像数据存在较大差异。本研究考虑了不同年份自然环境的差异性,选用2 a春季数据参与建模和验证,通过不断调整模型参数,建立最优反演模型,对春季土壤盐分进行反演,取得了良好的效果,但本次研究对象为非植被覆盖土,未能反演作物生长期的土壤盐分,所以下一步可以尝试研究作物光谱与土壤含盐量之间的关系。另外,试验区自然因素复杂多变,限制了土壤盐分的实时监测,因此反演方法在其他地区和其他季节的适用性还有待进一步研究。

4 结 论

(1) 棉田土壤主要为非盐化土和轻度盐化土,

总样本变异系数为0.67，呈中等变异性。光谱反射率与土壤盐渍化程度的关系表现为土壤盐渍化越重，光谱反射率越高。

（2）通过土壤盐分与OLI多光谱影像的波段、光谱指数进行相关性分析，发现b1、b2、b3、b4波段和SI1、SI2、SI3、SI4、S3、S4、S5盐分指数均通过显著性检验$P<0.01$，相关系数均达到0.4以上。

（3）所有模型中，基于全变量组建立的BPNN模型反演精度最高，建模集R^2为0.705，RMSE为1.133；验证集R^2为0.556，RMSE为1.409。综合分析建模和验证效果，选用基于全变量组的BPNN模型来反演春季棉田表层土壤含盐量。

（4）由反演结果可知，2019、2021年春季耕作区土壤主要为非盐化土，分别占耕作区总面积的55.55%和64.62%，其次为轻度盐化土，分别占44.31%和35.17%。越接近南北两侧荒地的耕作区土壤含盐量越高，东北部的小块耕地盐分偏高。2021年土壤盐渍化程度较2019年有所减轻。

参考文献（References）

[1] 梁萌, 米晓军, 李晨华, 等. 新疆准噶尔盆地未开垦盐碱土盐分与盐生植被多样性分析[J]. 干旱区地理, 2022, 45(1): 185−196. [Liang Meng, Mi Xiaojun, Li Chenhua, et al. Salinity characteristics and halophytic vegetation diversity of uncultivated saline-alkali soil in Junggar Basin, Xinjiang[J]. Arid Land Geography, 2022, 45(1): 185−196.]

[2] 孙亚楠, 李仙岳, 史海滨, 等. 基于多源数据融合的盐分遥感反演与季节差异性研究[J]. 农业机械学报, 2020, 51(6): 169−180. [Sun Ya'nan, Li Xianyue, Shi Haibin, et al. Remote sensing inversion of soil salinity and seasonal difference analysis based on multi-source data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 169−180.]

[3] 李晓明, 杨劲松, 杨奇勇. 基于电磁感应的南疆典型田块土壤盐分空间异质性研究[J]. 水土保持学报, 2011, 25(1): 167−170. [Li Xiaoming, Yang Jinsong, Yang Qiyong. Spatial heterogeneity of soil salinity in a typical field in south Xinjiang based on electromagnetic induction[J]. Journal of Soil and Water Conservation, 2011, 25(1): 167−170.]

[4] 李彪, 王耀强. 土壤盐渍化雷达反演模拟研究[J]. 干旱区资源与环境, 2015, 29(8): 180−184. [Li Biao, Wang Yaoqiang. Radar inversion and simulation of salty soil salinization[J]. Journal of Arid Land Resources and Environment, 2015, 29(8): 180−184.]

[5] Sreenivas K, Venkataratnam L, Rao P V N. Dielectric properties of salt-affected soils[J]. International Journal of Remote Sensing, 1995, 16(4): 641−649.

[6] Ding J L, Yu D L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. Geoderma, 2014, 235: 316−322.

[7] Rao B R M, Ravi Sankar T, Dwivedi R S, et al. Spectral behaviour of salt-affected soils[J]. International Journal of Remote Sensing, 1995, 16(12): 2125−2136.

[8] 奚雪, 赵庚星, 高鹏, 等. 基于Sentinel卫星及无人机多光谱的滨海冬小麦种植区土壤盐分反演研究——以黄三角垦利区为例[J]. 中国农业科学, 2020, 53(24): 5005−5016. [Xi Xue, Zhao Gengxing, Gao Peng, et al. Inversion of soil salinity in coastal winter wheat growing area based on sentinel satellite and unmanned aerial vehicle multi-spectrum: A case study in Kenli District of the Yellow River delta[J]. Scientia Agricultura Sinica, 2020, 53(24): 5005−5016.]

[9] 张素铭, 赵庚星, 王卓然, 等. 滨海盐渍区土壤盐分遥感反演及动态监测[J]. 农业资源与环境学报, 2018, 35(4): 349−358. [Zhang Suming, Zhao Gengxing, Wang Zhuoran, et al. Remote sensing inversion and dynamic monitoring of soil salt in coastal saline area[J]. Journal of Agricultural Resources and Environment, 2018, 35(4): 349−358.]

[10] 周晓红, 张飞, 张海威, 等. 艾比湖湿地自然保护区土壤盐分多光谱遥感反演模型[J]. 光谱学与光谱分析, 2019, 39(4): 1229−1235. [Zhou Xiaohong, Zhang Fei, Zhang Haiwei, et al. A study of soil salinity inversion based on multispectral remote sensing index in Ebinur Lake Wetland Nature Reserve[J]. Spectroscopy and Spectral Analysis, 2019, 39(4): 1229−1235.]

[11] 张晓光, 姜子璇, 孔繁昌. 滨海盐渍土可见近红外高光谱特征[J]. 遥感技术与应用, 2019, 34(4): 816−821. [Zhang Xiaoguang, Jiang Zixuan, Kong Fanchang. Hyperspectral characteristics of coastal saline soil with visible/near infrared spectroscopy[J]. Remote Sensing Technology and Application, 2019, 34(4): 816−821.]

[12] Zhang T T, Qi J G, Gao Y, et al. Detecting soil salinity with MODIS time series VI data[J]. Ecological Indicators, 2015, 52: 480−489.

[13] Alexakis D D, Daliakopoulos I N, Panagea I S, et al. Assessing soil salinity using WorldView-2 multispectral images in Timpaki, Crete, Greece[J]. Geocarto International, 2016, 33(4): 321−338.

[14] 张智韬, 魏广飞, 姚志华, 等. 基于无人机多光谱遥感的土壤含盐量反演模型研究[J]. 农业机械学报, 2019, 50(12): 151−160. [Zhang Zhitao, Wei Guangfei, Yao Zhihua, et al. Soil salt inversion model based on UAV multispectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 151−160.]

[15] 冯雪力, 刘全明. 基于多源遥感协同反演的区域性土壤盐渍化监测[J]. 农业机械学报, 2018, 49(7): 127−133. [Feng Xueli, Liu Quanming. Regional soil salinity monitoring based on multi-source collaborative remote sensing data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(7): 127−133.]

[16] 刘恩, 王军涛, 常步辉, 等. 小开河引黄灌区土壤盐渍化定量遥感反演[J]. 中国农村水利水电, 2019(12): 20−24. [Liu En, Wang

Juntao, Chang Buhui, et al. Quantitative remote sensing inversion of soil salinization in Xiaokaihe Yellow River irrigation district[J]. China Rural Water and Hydropower, 2019(12): 20–24.]

[17] 张智韬, 杜瑜燕, 劳聪聪, 等. 基于雷达遥感的不同深度土壤含盐量反演模型[J]. 农业机械学报, 2020, 51(10): 243–251. [Zhang Zhitao, Du Yuyan, Lao Congcong, et al. Inversion model of soil salt content in different depths based on radar remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 243–251.]

[18] 何宝忠, 丁建丽, 刘博华, 等. 渭库绿洲土壤盐渍化时空变化特征[J]. 林业科学, 2019, 55(9): 185–196. [He Baozhong, Ding Jianli, Liu Bohua, et al. Spatiotemporal variation of soil salinization in Weigan-Kuqa River Delta Oasis[J]. Scientia Silvae Sinicae, 2019, 55(9): 185–196.]

[19] 阿尔达克·克里木, 塔西甫拉提·特依拜, 张东, 等. 基于高光谱的ASTER影像土壤盐分模型校正及验证[J]. 农业工程学报, 2016, 32(12): 144–150. [Kelimu Ardak, Tiyip Tashpolat, Zhang Dong, et al. Calibration and validation of soil salinity estimation model based on measured hyperspectral and ASTER image[J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32 (12): 144–150.]

[20] 丁邦新, 白云岗, 柴仲平, 等. 塔里木河下游绿洲灌区土壤盐渍化特征及季节性变化规律[J]. 水土保持通报, 2020, 40(2): 77–84. [Ding Bangxin, Bai Yungang, Chai Zhongping, et al. Soil salinization characteristics and its seasonal variation in oasis irrigation district of lower reaches of Tarim River[J]. Bulletin of Soil and Water Conservation, 2020, 40(2): 77–84.]

[21] 叶勤, 姜雪芹, 李西灿, 等. 基于高光谱数据的土壤有机质含量反演模型比较[J]. 农业机械学报, 2017, 48(3): 164–172. [Ye Qin, Jiang Xueqin, Li Xican, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 164–172.]

[22] Marshall F E, Smith D T, Nickerson D M. Empirical tools for simulating salinity in the estuaries in Everglades National Park, Florida [J]. Estuarine Coastal & Shelf Science, 2011, 95(4): 377–387.

[23] 冯娟, 丁建丽, 杨爱霞, 等. 干旱区土壤盐渍化信息遥感建模[J]. 干旱地区农业研究, 2018, 36(1): 266–273. [Feng Juan, Ding Jianli, Yang Aixia, et al. Remote sensing modeling of soil salinization information in arid areas[J]. Agricultural Research in the Arid Areas, 2018, 36(1): 266–273.]

[24] 吴忠强, 毛志华, 王正, 等. 基于极限学习机的浅海水深遥感反演研究[J]. 海洋测绘, 2019, 39(3): 11–15. [Wu Zhongqiang, Mao Zhihua, Wang Zheng, et al. Research on remote sensing inversion of shallow water depth based on extreme learning machine[J]. Hydrographic Surveying and Charting, 2019, 39(3): 11–15.]

[25] 陆军胜, 陈绍民, 黄文敏, 等. 采用SE_{PLS}_ELM模型估算夏玉米地上部生物量和叶面积指数[J]. 农业工程学报, 2021, 37(18): 128–135. [Lu Junsheng, Chen Shaomin, Huang Wenmin, et al. Estimation of aboveground biomass and leaf area index of summer maize using SE_{PLS}_ELM model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(18): 128–135.]

[26] 姜雯, 吴陈. 基于改进粒子群算法的支持向量机遥感影像分类[J]. 江苏科技大学学报(自然科学版), 2020, 34(5): 66–72. [Jiang Wen, Wu Chen. Remote sensing image classification by support vector machine based on improved particle swarm optimization[J]. Journal of Jiangsu University of Science and Technology (Natural Science Edition), 2020, 34(5): 66–72.]

[27] 赵洪莹, 舒清态, 罗文秀, 等. 基于Landsat 8-OLI的高山松叶面积指数采样尺度优化分析[J]. 西南林业大学学报(自然科学版), 2021, 41(5): 114–120. [Zhao Hongying, Shu Qingtai, Luo Wenxiu, et al. Optimization analysis of sampling scale of *Pinus densata* leaf area index based on Landsat 8-OLI[J]. Journal of Southwest Forestry University (Natural Sciences Edition), 2021, 41(5): 114–120.]

[28] 李柏年, 吴礼斌. MATLAB数据分析方法[M]. 北京: 机械工业出版社, 2012: 170–171. [Li Bainian, Wu Libin. MATLAB data analysis methods[M]. Beijing: China Machine Press, 2012: 170–171.]

[29] 张录, 张芳, 熊黑钢, 等. 不同季节强碱土土壤呼吸影响因子分析与模型预测[J]. 干旱地区农业研究, 2017, 35(1): 71–78. [Zhang Lu, Zhang Fang, Xiong Heigang, et al. Impact factor analysis and model prediction of strong alkaline soil respiration in different seasons[J]. Agricultural Research in the Arid Areas, 2017, 35(1): 71–78.]

[30] 新疆维吾尔自治区农业厅, 新疆土壤普查办公室. 新疆土壤[M]. 北京: 科学出版社, 1996: 151–521. [Xinjiang Department of Agriculture, Xinjiang Soil Survey Office. The soil of Xinjiang[M]. Beijing: Science Press, 1996: 151–521.]

[31] Peck A J. Note on the role of a shallow aquifer in dryland salinity [J]. Soil Research, 1978, 16(2): 237–240.

[32] 柳菲, 陈沛源, 于海超, 等. 民勤绿洲不同土地利用类型下土壤水盐的空间分布特征分析[J]. 干旱区地理, 2020, 43(2): 406–414. [Liu Fei, Chen Peiyuan, Yu Haichao, et al. Spatial distribution characteristics of soil water and salt under different land use types in Minqin Oasis[J]. Arid Land Geography, 2020, 43(2): 406–414.]

[33] 姜凌, 李佩成, 胡安焱, 等. 干旱区绿洲土壤盐渍化分析评价[J]. 干旱区地理, 2009, 32(2): 234–239. [Jiang Ling, Li Peicheng, Hu Anyan, et al. Analysis and evaluation of soil salinization in oasis of arid region[J]. Arid Land Geography, 2009, 32(2): 234–239.]

[34] 王雪梅, 周晓红. 渭干河—库车河三角洲绿洲棉田土壤盐分估算及遥感反演[J]. 干旱地区农业研究, 2018, 36(6): 250–254, 262. [Wang Xuemei, Zhou Xiaohong. Estimation and inversion modeling of salinity of cotton field soil using remote sensing in the delta oasis of Weigan and Kuqa Rivers[J]. Agricultural Research in the Arid Areas, 2018, 36(6): 250–254, 262.]

[35] 边玲玲, 王卷乐, 郭兵, 等. 基于特征空间的黄河三角洲垦利县土壤盐分遥感提取[J]. 遥感技术与应用, 2020, 35(1): 211–218. [Bian Lingling, Wang Juanle, Guo Bing, et al. Remote sensing extraction of soil salinity in Yellow River delta Kenli County based on feature space[J]. Remote Sensing Technology and Application, 2020, 35(1): 211–218.]

[36] 贾萍萍, 张俊华, 孙媛, 等. 基于实测高光谱和Landsat 8 OLI影像的土壤盐化和碱化程度反演研究[J]. 土壤通报, 2020, 51(3): 511−520. [Jia Pingping, Zhang Junhua, Sun Yuan, et al. Inversion of soil salinity and pH degree based on measured hyper-spectral and Landsat 8 OLI image[J]. Chinese Journal of Soil Science, 2020, 51(3): 511−520.]

[37] 杨小虎, 罗艳琴, 杨海昌, 等. 玛纳斯河流域绿洲农田土壤盐分反演及空间分布特征[J]. 干旱区资源与环境, 2021, 35(2): 156−161. [Yang Xiaohu, Luo Yanqin, Yang Haichang, et al. Soil salinity retrieval and spatial distribution of oasis farmland in Manasi River Basin[J]. Journal of Arid Land Resources and Environment, 2021, 35(2): 156−161.]

Inversion and validation of soil salinity based on multispectral remote sensing in typical oasis cotton field in spring

LIU Xuhui[1,2], BAI Yungang[2], CHAI Zhongping[1], ZHANG Jianghui[2], DING Bangxin[2,3], JIANG Zhu[2]

(1. College of Resources and Environment, Xinjiang Agricultural University, Urumqi 830052, Xinjiang, China; 2. Xinjiang Research Institute of Water Resources and Hydropower, Urumqi 830049, Xinjiang, China; 3. College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling 712100, Shaanxi, China)

Abstract: This study aimed to explore an effective method for extracting soil salinity from cotton fields in oasis and determine the characteristics and spatial distribution of soil salinization on a regional scale to provide reference for soil salinization control. The 31st Regiment of the 2nd Division of Xinjiang Production and Construction Corps was taken as the research area; Landsat 8 OLI multispectral images and field measurements of soil salinity in spring 2019 and 2021 were taken as data sources; and the band group, spectral index group, and total variable group were taken as the model input variable group. Multiple stepwise regression (MSR), partial least squares regression, extreme learning machine, support vector machine, and back propagation neural network (BPNN) were used to construct a remote sensing inversion model of soil salt based on the three input variable groups. Precision evaluation was conducted, and the effects of input variables and modeling methods on model accuracy were explored. The best inversion model of soil salt in spring was determined through comparison and quantitatively inverted surface soil salt content. Results showed that (1) the study area was mainly composed of non-salinized soil and slightly salinized soil, and the coefficient of variation of total samples was 0.67, implying moderate variability. The relationship between spectral reflectance and soil salinization is as follows: serious soil salinization leads to great spectral reflectance. (2) Significance tests were conducted on coastal band(b1), blue (b2), green(b3), and red(b4) and the salinity indices of SI1, SI2, SI3, SI4, S3, S4, and S5 (P<0.01). The obtained correlation coefficients were all above 0.4, which can represent soil salinity to a certain extent. (3) Among the six linear regression models, the MSR model established on the band group had the best inversion effect, and the BPNN inversion model established on the full variable group had the highest accuracy. (4) According to the inversion results, the soil in spring 2019 and 2021 was mainly nonsalinized soil accounting for 55.55% and 64.62% of the total tillage area, respectively, followed by mild salinized soil accounting for 44.31% and 35.17%, respectively. Soil salinization in 2021 was reduced compared with that in 2019.

Key words: multispectral remote sensing inversion; soil salinity; spectral reflectance; variable group; machine learning